

Deep Learning for Detecting Abandoned Dogs

Oluwakemi Akinwehinmi^{1,2} , Alberto Tena¹ , Francisco Javier Mora² , Francesc Solsona¹ , Pedro Arnau del Amo³ 

¹ Departament d'Enginyeria Informàtica i Disseny Digital, Universitat de Lleida, Carrer de Jaume II, 69 25001 Lleida (Spain)

² Centre Internacional de Mètodes Numèrics en Enginyeria (CIMNE), CIMNE -Edifici C1 Campus Nord UPC C/ Gran Capità, S/N, Barcelona, 08034 (Spain)

³ Findspo, Calle de Cós, 20, Sallent, Barcelona 08650 (Spain)

* Corresponding author: alberto.tena@udl.cat

Received 1 August 2024 | Accepted 19 December 2025 | Early Access 5 March 2026



ABSTRACT

This research paper presents a methodology consisting of an algorithm and a workflow for finding abandoned dogs in natural surroundings. We propose a temporal-contextual methodology for identifying abandoned pets in public areas. This involves employing a temporal rule alongside context-sensitive object identification, whereby dog bounding boxes are deliberately expanded to encompass proximate visual indicators suggestive of abandonment. The proposed approach uses object detection techniques, trajectory analysis, and image segmentation to quickly differentiate between abandoned and owned dogs. The research addresses key challenges, such as data scarcity and the complexity of distinguishing between abandoned and accompanied dogs. To address the issue of sufficient and adequate training corpus to identify an abandoned dog, the research article employs single-channel image augmentation methods that improve model recall and precision by 4%. Several object detection algorithms were evaluated, and our findings indicate that single-stage detectors like YOLOv8 achieved a better trade-off between classification performance and speed for detecting abandoned dogs compared to multi-stage detectors like Faster R-CNN, reaching a mean average precision up to 86% and an inference time of 0.3 ms per frame. This research study contributes to animal welfare, biodiversity conservation, and public safety by providing a scalable solution for monitoring abandoned animals in diverse environments. The findings demonstrate the impact of object detection techniques on improving the generalization of deep learning models for real-world applications.

KEYWORDS

Activity Recognition, Algorithms, Anomaly Detection, Computer Vision, Deep Learning, Edge Computing, Machine Learning, Machine Vision.

DOI: [10.9781/ijimai.2026.2222](https://doi.org/10.9781/ijimai.2026.2222)

I. INTRODUCTION

INVASIVE species which can be referred to as abandoned animals, in natural environments pose a significant risk to biodiversity, human health, safety, and environmental stability. These abandoned animals compete with local species for resources, alter predator-prey dynamics, and may spread illnesses. Traditional methods of finding abandoned pets, which generally involve physical searches or rely on public reports, are often hampered by limitations in manpower and the vast areas needing coverage.

Fortunately, advancements in computer vision, particularly deep learning, offers a promising solution to address these limitations. Deep learning algorithms have made significant progress in object detection tasks in recent years [1], [2].

From an ecological perspective, abandoned dogs function as invasive agents, disrupting native species dynamics via predation, interspecific competition, and disease transmission, thereby presenting

a significant danger to the integrity of protected natural areas [3]. The increasing population of abandoned and free-roaming dogs in urban and periurban areas poses substantial problems for biodiversity conservation, public health, and animal welfare. More than 200 million stray dogs worldwide are predicted to be not included in current urban planning, veterinary infrastructure, and waste management systems [4]. In numerous low- and middle-income nations, these animals significantly facilitate the transmission of zoonotic diseases, particularly rabies, which results in approximately 59,000 human deaths annually, mainly due to dog bites in regions with inadequate animal control measures [5].

Existing deep learning algorithms for object detection, such as YOLO (You Only Look Once) models (YOLOv5 [6], YOLOv6 [7], and YOLOv8), Fast and Faster R-CNN (Faster Region-based Convolutional Neural Network) [8], [9], and the Facebook's Detectron2 framework [10], [11], are state-of-the-art deep learning algorithms designed to identify and classify objects in images. YOLO models, for instance, prioritize speed and efficiency by processing images in a single pass,

Please cite this article as:

O. Akinwehinmi, A. Tena, F. J. Mora, F. Solsona, P. Arnau del Amo. Deep Learning for Detecting Abandoned Dogs, International Journal of Interactive Multimedia and Artificial Intelligence, (2026), <http://doi.org/10.9781/ijimai.2026.2222>

making them ideal for real-time applications. In contrast, Fast and Faster R-CNN, along with Facebook's Detectron2, employ region-based methods that enhance classification performance in detecting multiple objects, albeit at the cost of increased computational requirements. A detailed comparison of these algorithms, including their pros and cons, would be essential to determine the most effective approach to detecting abandoned dogs.

These state-of-the-art algorithms depend on three important components: powerful computational hardware like GPUs, effective algorithms, and vast datasets of training images. However, gathering and labelling huge datasets of abandoned dogs in natural settings can be costly, time-consuming, and impracticable.

Identifying abandoned dogs in natural and urban environments are complex due to the vast variety of appearances and unpredictable situations. To overcome these issues, we offer a unique methodology that uses a trajectory analysis method to track dog movement patterns, as well as deep learning for object detection. To improve context awareness, we propose a simple yet effective modification: enlarging bounding boxes around detected dogs. This contextual extension helps capture nearby visual cues such as absence of humans which are critical for abandonment detection.

It was inspired by these methodologies [1] employed to identify abandoned luggage in airports, where dynamic, time-based categorization proved more accurate than static classification [12]. Our approach significantly impacts requirements by enabling real-time categorization and reaction [13], ensuring prompt identification and response. Additionally, the system is designed to adapt to various environmental conditions, maintaining reliability and trustworthiness across different scenarios.

Leveraging the advancements in computer vision algorithms, especially those utilizing deep learning techniques, this study examines the efficacy of different strategies for detecting abandoned dogs. We explore whether deep learning-based computer vision systems can accurately detect abandoned pets despite visual changes and contextual variations. Additionally, we examine the effectiveness of data augmentation techniques, such as single-channel augmentation, in enhancing model performance by mitigating data scarcity and the inherent unpredictability of real-world settings.

A significant problem in applying deep learning to real-world applications, such as abandoned dog identification, is getting high model performance with little datasets. Model generalization refers to a model's capacity to properly recognize abandoned dogs in new conditions that it has not experienced during training. Models with weak generalization tend to overfit the training data, performing well on certain photos but failing to detect abandoned dogs in previously encountered settings. This emphasizes the need for novel approaches to the problem of insufficient training data.

Data augmentation techniques [14], including single-channel image augmentation, expand the training datasets by applying transformations like rotations, translations, and color adjustments [15]. This leads to enhanced model generalization and performance.

A. Research Gaps and Motivation

The need to better understand human presence in natural environments and its impact on issues such as pet abandonment and biodiversity monitoring served as the primary motivation for this research. These insights are essential for informing targeted interventions and advancing the use of computer vision in ecological contexts. This motivation aligns with the goals of our research project Portal for the Integration of Knowledge for a Sustainable Ecosystems and Land Management also known as PIKSEL [16], funded by the Catalan Government, which seeks to monitor human activity in natural areas to support sustainable land management strategies.

Secondly, object detection models, including YOLO models and Faster R-CNN, predominantly depend on static image classification and fail to integrate contextual awareness or temporal data when differentiating between visually analogous situations, such as a dog briefly unattended versus one that has been genuinely abandoned. These constraints impede their capacity to precisely discern dynamic, real-world behaviors in outside environments. Our proposed methodology addresses this gap by introducing a context-aware framework that integrates both tight and loose bounding boxes with trajectory-based behavioral analysis.

B. Research Questions

The context and gaps identified led us to formulate the following research questions:

- Can deep learning-based computer vision algorithms accurately distinguish between dogs with and without human companions in real-world settings (natural and urban), despite differences in appearance and surroundings?
- What is the speed and classification performance of Abandoned dog detection method in real-time settings?
- How effectively do alternative object detection algorithms (YOLO models) perform in real-time detection of abandoned dogs using 'loose' bounding boxes that incorporate contextual information?
- How do existing object detection algorithms (e.g., YOLO models and Faster R-CNN) perform in detecting and classifying abandoned dogs?
- Do data augmentation approaches, including single-channel image augmentation, improve the performance of deep learning models for detecting abandoned dogs?

C. Hypothesis

1. Computer vision algorithms based on deep learning can detect and classify abandoned dogs with high classification performance in a variety of natural environments
2. Adaptive image data augmentation, particularly single-channel augmentation, improves model performance by mitigating data scarcity and variability.
3. For detecting abandoned dogs in natural surroundings, single-stage detection algorithms like YOLO models will outperform multi-stage algorithms such as Faster R-CNN in terms of speed while achieving comparable classification performance.

D. Main Contributions

This research paper provides three main contributions:

1. An innovative methodology that employs context-aware detection through "loose" bounding boxes to enhance classification performance in natural settings, based on a novel algorithm for trajectory analysis specifically tailored to dog movement. This approach adapts to diverse conditions by leveraging advanced object detection models and incorporating both contextual and visual information for effective dog classification [17].
2. A suitable workflow architecture for our detection algorithm that enables real-time detection of dog abandonment by analyzing consecutive frames for the presence of nearby owners. We also identified the most suitable detection models for this endeavor.
3. Adaptive and automated image augmentation techniques [18], [19] are introduced to enhance the algorithm's performance. The existing method "Adaptive Image Data Augmentation for Deep Learning Models" is leveraged to address data size constraints by augmenting small datasets, with a focus on single-channel image augmentation.

II. LITERATURE REVIEW

A. Deep Learning Approaches to Abandoned Object Detection

Current methods for detecting abandoned items, such as abandoned luggage in public places like airports or cars [20], are based on object tracking within the scene. However, these techniques frequently fail to distinguish between people and objects, making them ineffective for locating abandoned dogs [21]. Furthermore, current research focuses on static objects, while abandoned dogs are dynamic and mobile [22].

Deep learning has demonstrated potentials in the aspect of object identification, with algorithms analyzing video frames [23] to identify abandoned pets based on cues such as absence of a nearby person [24]. Static image datasets for detection has limits. A dog lying down or briefly separated from its owner may be misidentified as abandoned in a single shot.

By tracking dog movement patterns over time in video data, we can gain valuable insights into their behavior. Analyzing factors such as prolonged inactivity or movement patterns indicating a lost or disoriented dog can considerably increase the classification performance of abandoned dog identification.

B. A Deep Learning and Trajectory Analysis Framework for Abandoned Dog Detection

YOLO models are known as single-stage detectors which are noted for their high speed and efficiency, making them ideal for real-time applications [7]. YOLOv6 and YOLOv8 in particular, has been created with an emphasis on industrial applications, providing better performance and resilience.

Two-stage detectors such as Faster R-CNN have long been renowned for their higher classification performance, albeit at the cost of slower processing rates. These models create region suggestions first, then categorize and refine the bounding boxes, resulting in more accurate item identification but with greater processing costs.

Similar strategies to the one proposed in this study, which emulates human cognitive processes when assessing a dog's abandonment status [25], have been followed by other works [23], [26], proving effective in related fields such as animal monitoring and urban street surveillance, with high classification performance in object detection and contextual understanding.

Deep learning has potential in finding abandoned pets, particularly dogs [23]. However, there are significant problems in applying these strategies to this particular task. Traditional object recognition datasets often lack the diversity and quality necessary to build efficient deep learning models [27].

Recent advances in data augmentation techniques [15], [28] offer a solution ahead. Past research strategies increase datasets by altering existing images, hence enhancing model performance and usefulness. The introduction of the ImageNet [29] database and the LSVRC (Large Scale Visual Recognition Challenge) contributed to demonstrate the efficiency of data augmentation for deep learning tasks [15], [28]. Random cropping, translation, flipping and RGB (Red, Green Blue) channel manipulation [30] have all been demonstrated to increase image classification performance [18], [19], [31]–[34].

C. Challenges in Detecting Abandoned Dogs in Natural Settings

Recent research has investigated modern augmentation techniques to overcome these limitations [31]. To expand datasets, methods such as Set Pattern [33], [35], Augmentation [28] and Mix-up [32] have been proposed. However, these techniques can be overly aggressive, resulting in unrealistic distortions and potential biases [15], [35]. For example, while rotation augmentation may be useful for datasets such as CIFAR-10 [36], it can be detrimental for tasks such as abandoned dog

detection by confusing dog breeds or making it difficult to distinguish a dog lying down from an abandoned object, particularly in natural settings with uneven terrain [21].

To address these issues, researchers developed techniques such as AutoAugment [18] and Fast AutoAugment [19], which use reinforcement learning to optimize augmentation strategies for specific tasks [31], [37], [38]. However, performance inconsistencies remain an issue [35].

D. Data Limitations in Abandoned Dog Detection

Past research works have addressed the challenges associated with datasets, data quality and quantity. In the field of biodiversity, traditional datasets for general object recognition frequently lack the necessary diversity in dog breeds, poses and lighting conditions unique to natural environments, as well as the absence of a human companion. All of them are critical for robust model training in abandoned dog detection [27]. Although the authors in [28], [39] introduced a new annotated dataset featuring various dog breeds in diverse scenarios, further progress is necessary to overcome data limitations and enhance the effectiveness of deep learning models for this specific task.

E. Comparison of Previous Studies on Abandoned Dog Detection

Zhao et al. [40] presents a deep learning-assisted system for real-time monitoring of search and rescue dogs' actions, audio signals, and positions. This framework uses wearable sensors to collect complete behavioural data, which is then analysed using powerful deep learning algorithms to recognise unique behaviours. Such methodologies could be adapted for abandoned dog detection by observing their movements and behaviors in varied environments.

Azizi and Zaman [39] investigate the application of deep learning to enhance the process of locating lost pets, specifically cats and dogs. The authors use transfer learning techniques across several Convolutional Neural Networks (CNNs) to effectively classify species and identify individuals. Their approach includes important procedures including data preparation, species categorisation, face and body detection, alignment, and identification.

Khan et al. [41] performed a comprehensive assessment of real-time dog identification systems that used artificial intelligence technologies. Their research emphasises important advances in deep learning approaches, exploring numerous algorithms and frameworks to address difficulties linked to detection classification performance and processing performance across diverse settings. The authors emphasize the critical role of real-time applications in promoting animal welfare and effectively managing stray populations.

Ruiz-Chavez et al. [26] analyze the socio-public health impacts of dog abandonment in Quito, Ecuador, employing georeferenced multi-agent systems to simulate the dynamics of abandoned dog populations. Their study is on the relationship between geographic and social elements and the issue of abandonment, encouraging the use of simulation tools for optimal policy planning. The study emphasises the significance of knowing the broader consequences of dog abandonment in order to guide focused treatments.

III. METHODOLOGY

In this section, we present our innovative methodology for the real-time identification of invasive species in natural contexts, specifically tailored for detecting abandoned dogs. This methodology is grounded in two main components: the detection algorithm and the workflow for identifying abandoned dogs in natural environments. Both components are explained in detail throughout this section. The datasets and codes used for the experiments described in Subsection

F have been made publicly available on Kaggle [42] and GitHub [43], respectively, to support reproducibility and encourage further research in this area.

A. Datasets

As a primary source of data, we compiled a diverse dataset of 500 images and 20 videos of dogs with and without owners at various natural spaces of Barcelona. This corpus helped the model distinguish between abandoned and not abandoned dogs. We also used a secondary source of data (Kaggle COCO datasets) for training our model, we augmented a Kaggle animal image dataset (17,500 images) to create a new training set with 35,000 single-channel images per class (Red, Blue, Green: 3 channels total). For this endeavor, we used the single-color channel augmentation technique, which is explained in detail in the following subsection.

B. Image Augmentation With Single-Channel Color Space Manipulation

In real-world settings, accurately detecting abandoned dogs can be challenging due to factors like variations in lighting, pose, and background clutter. To address this, we leverage existing image data augmentation techniques. This approach artificially expands our training dataset by generating variations of existing dog images. By incorporating these augmented images, our system is exposed to a wider range of scenarios and learns more robust features. This, in turn, improves the generalization of our dog detection and classification algorithm, allowing it to perform more accurately in real-world conditions.

We leverage a single-channel color space manipulation technique for on-the-fly image augmentation. Inspired by insights from previous studies [15], a CNN is employed to pre-process the initial image dataset. CNN extracts feature from the images in each video frame, resizes the images to a pre-defined size, facilitating normalization. Normalization is a common practice in deep learning that scales pixel values to a specific range (often between 0 and 1) to improve training stability and convergence.

The images in the video frames are initially converted from the BGR (Blue, Green, Red) colour structure, which is widely used in computer vision libraries such as OpenCV, to the RGB colour space. The alteration is critical because most CNN patterns are built to handle images in RGB mode. To enhance model performance, we modify the images from RGB to HSV (Hue, Saturation, Value) colour structure. This reduces the model's sensitivity to differences in lighting and object positioning in the images. A class-based framework for image augmentation and preprocessing ensures that data is correctly designed for deep learning models.

Splitting the image into different colour channels (HSV) is an important part of the augmentation approach. This is accomplished using OpenCV's `cv2.split` function. As a result of channel separation, the number of channels in the improved datasets grows. To maintain unique identifiers for each augmented image, we propose attaching an additional value channel containing a unique identifier for each augmented image.

Fig. 1 and Fig. 2 illustrate the single-color channel augmentation technique used to augment the dataset. Fig. 1 shows the original image, while Fig. 2 presents the result of applying this technique. Although no apparent differences are visible, image augmentation transforms existing images into new ones through techniques such as cropping, blurring, and color adjustments. While augmentation is beneficial, it must be balanced with other techniques to prevent the model from overfitting to the training data.



Fig. 1. Initial Image.



Fig. 2. Augmented Image.

We evaluated the effectiveness of this image augmentation technique on model performance [28]. The details about the experiment we conducted are provided in the Subsection F addressing the recommendation by Connor and Taghi [15] to explore color space augmentation across datasets in image recognition tasks.

C. Method for the Detection of Abandoned Dogs

In this subsection, we describe our method for identifying abandoned dogs. Our approach integrates image segmentation to enhance object distinction and trajectory analysis, incorporating both optical flow and dwell time, to effectively monitor movement patterns and differentiate abandoned dogs from those with owners.

Before detailing the technical workflow employed in our detection system, we first describe the behavioral assumptions and interaction patterns that guided the design of our methodology.

1. Assumptions and Behavioral Characteristics of Abandoned Dogs

Our research methodology is based on several key assumptions of the behavior and characteristics of abandoned dogs, which are detailed as follows:

- Abandoned dogs are unchained and roam freely, without physical confinement. [23] investigates smartphone and GPS technology for monitoring free-roaming dog populations, implying abandoned dogs might exhibit similar behavior.
- Abandoned dogs tend to move in unpredictable patterns, unlike stationary objects. In [24], an important reference to stray dogs suggests they might not exhibit the same predictable behavior as leashed or working dogs. Authors in [35] also focused on detecting suspicious background changes in video surveillance, which could be applicable to the unpredictable movements of abandoned dogs.

- Abandoned dogs are unlikely to have human contact for long periods of time. Authors in [25] investigated spotting strays and abandoned pets, which implies a lack of regular human contact.
- Usually, abandoned dogs are smaller than humans. This can also be likened to abandoned cats, rabbits etc as they are similar in size characteristics.

These assumptions provide the foundation for creating the Algorithm 1 capable of detecting abandoned dogs in natural settings.

Algorithm 1. Process for Abandoned Dog Detection

```

1: Step 1: Dataset Preprocessing
2: for each video  $V$  in dataset  $D$  do
3:    $F \leftarrow \text{DecomposeIntoFrames}(V)$ 
4:   for each frame  $f$  in  $F$  do
5:      $f_{\text{normalized}} \leftarrow \text{Normalize}(f)$ 
6:      $f_{\text{augmented}} \leftarrow (f_{\text{normalized}})$ 
7:   end for
8: end for

9: Step 2: Object Detection
10: for each frame  $f$  in  $F$  do
11:    $objects \leftarrow \text{DetectObjects}(f)$ 
12:    $D_f \leftarrow \text{IdentifyInstances}(objects, "dog")$ 
13:    $P_f \leftarrow \text{IdentifyInstances}(objects, "person")$ 
14: end for

15: Step 3: Dog-Human Interaction Analysis
16: for each frame  $f$  with  $D_f$  and  $P_f$  do
17:   if  $D_f$  and  $P_f$  are present then
18:      $S \leftarrow \text{CalculateProximity}(D_f, P_f)$ 
19:      $\text{RecordSpatialConnection}(S, t)$ 
20:   end if
21: end for

22: Step 4: Time-Based Abandoned Dog Classification
23:  $T \leftarrow \text{time\_threshold}$ 
24: Initialize tracking system  $T_r$  for  $D_f$ 
25: for each dog  $d$  in  $D_f$  do
26:   Track  $d$  for  $T$  minutes
27:   while tracking  $d$  do
28:     if  $P_f$  is within proximity  $S$  of  $d$  then
29:       Mark  $d$  as "Not Abandoned"
30:     else if no  $P_f$  within proximity  $S$  of  $d$  after  $T$  minutes then
31:       Proceed to Step 5
32:     end if
33:   end while
34: end for

35: Step 5: Classifying and Responding to Abandoned Dogs
36: for each  $d$  after  $T$  minutes do
37:   if no  $P_f$  detected then
38:     Classify  $d$  as "Abandoned"
39:      $\text{TriggerResponseAction}(d)$ 
40:   else
41:     Mark  $d$  as "Not Abandoned"
42:   end if
43: end for

```

2. Image Segmentation

Image segmentation is applied to separate the dogs and humans from its surroundings. Image segmentation is the process of splitting an image dataset into several segments or areas to facilitate analysis, which is commonly used for techniques such as object identification, recognition, and classification. We started with image preprocessing, which included noise removal and normalization. Noise removal and normalization were followed by boundary extraction and segmentation. We followed a sequence of processes that efficiently separated dogs and humans from the the background. We started with preprocessing technique to eliminate noise and improve image quality, then normalised the pixel intensities across images. For boundary detection, we used edge detection methods to highlight edges in the images, which helped us separate items of interest from the background. Finally, we employed clustering-based algorithm to segment the image and distinguish between dogs and humans.

Fig. 3 illustrates the segmentation method used to distinguish between abandoned and accompanied dogs. The top part of the figure shows a dog enclosed within a red bounding box, accompanied by a red segmentation mask on the right that accurately isolates the silhouette of the dog. In contrast, the bottom left image depicts a person walking a dog, enclosing the person within a green bounding box.

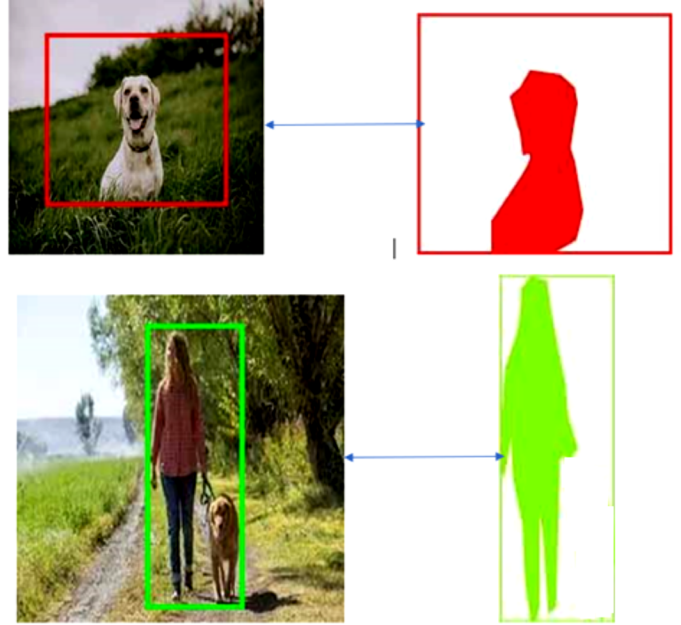


Fig. 3. Image segmentation of dogs and humans in Natural Space.

3. Trajectory Analysis Using Optical Flow

Detecting the dogs and persons from the background clutter, tracking algorithms can determine their movements more precisely.

Dogs and person's interaction are obtained from their segmentation and trajectory analysis. The intersection of the trajectories of dogs and people will determine whether the dog has contacted its owner, thus identifying abandoned dogs.

Trajectory analysis is the study of an object's path over time, it provides insights into motion patterns, behavior prediction, and anomaly identification. Trajectory analysis combines two fundamental techniques: optical flow and dwell duration. Optical flow calculates an object's velocity and direction by examining apparent motion patterns in an image. We employed the Lucas-Kanade approach [44] for minor movements and the Horn-Schunck method [45] for global motion estimate. The Lucas-Kanade Approach analyzes tiny image patches

surrounding a specified location, assuming that the motion within such patches is constant, while the Horn-Schunck Method estimates total motion throughout the whole image which takes into account smoothness requirements, and prioritizes solutions with motion vectors that are comparable to those of neighboring pixels.

4. Trajectory Analysis Using Dwell Time

Dwell time measures how long an object stays in a given position. This data is critical for understanding object behavior, such as determining regions of interest or persistent activity. To keep the detected dog's identification consistent over frames, we applied a DeepSORT tracking method [46], [47]. This tracking allowed us to continually monitor the dog's movement and guarantee that it was correctly identified as it passed across the frame.

We identified "regions of interest" inside the frame to act as thresholds for monitoring dwell duration without human companionship. By establishing this, we were able to focus on the dog's behavior which is critical for spotting extended or wandering stays that indicate probable abandonment.

We preset a time-frame the identified dog spent in the designated region over numerous frames. Each time the object was seen in the selected zone, the dwell duration increased, providing a cumulative estimate of its presence in that region.

We specified a dwell time threshold, after which an alarm would be triggered. This threshold served as an indication, implying that the dog may be abandoned if it stayed in the location after the stated time limit.

Fig. 4 illustrates how our model, using pattern recognition, recognises, classify and predicts that the person is actively strolling with more than one dog. The bounding box including the human and the dogs is labelled 'not abandoned' with a confidence score of 0.80. This label indicates that the model has high confidence that these dogs are escorted rather than abandoned.



Fig. 4. Status Classification.

D. Workflow and Pseudo-Code for the Abandoned Dog Classification Algorithm

1. Workflow for the Abandoned Dog Classification Algorithm

The workflow for our abandoned dog classification algorithm is structured into five interconnected stages, which are explained next, guiding the process from initial detection to final classification. Each stage is designed to systematically address specific challenges in real-time monitoring, ensuring accurate identification through a streamlined and adaptable sequence.

Fig. 5 provides a schematic illustration of the workflow implemented for detecting abandoned dogs in video frames. The process begins with dataset framing (Step 1) and object detection (Step 2), where dogs and persons are identified and their proximity is assessed. Then, the detected dog is analyzed in the frame (Step 3), followed by a

temporal evaluation of its presence over a fixed period (Step 4). If a person is detected near the dog, the system classifies the situation as "Unabandoned Dog". Otherwise, if the dog remains alone for more than the fixed period, it is labeled as an "Abandoned Dog" (Step 5).

Each step of this workflow is described in detail below, where the specific criteria and methods used at each stage are thoroughly explained.

1. Step 1: Dataset Preprocessing.

The video datasets are broken down into video frames. Previous research analyzed images without considering temporal changes. Our technique advances from static images to videos and real-world recordings, incorporating temporal factors to enhance the detection classification performance of abandoned pets. By breaking videos into frames, our method supports parallel processing, thereby improving real-time capabilities. Additionally, our approach adapts to varying environmental conditions, such as changes in illumination and weather, through preprocessing techniques like normalization and augmentation.

2. Step 2: Object detection.

Object detection techniques are used to locate and categorize objects of interest inside each frame. This involves detecting dog and human instances.

3. Step 3. Dog-Human Interaction Analysis.

The closeness between detected dogs and person instances is used to estimate their spatial connection. In this step, we assess the spatial relationship between detected dogs and persons in each frame by calculating their proximity. Whenever both a dog and a person appear in a frame, we calculate the spatial distance between detected dogs and persons in each frame to estimate their proximity. For every frame where both objects appear, the system records this distance along with a timestamp (t), allowing us to build a temporal sequence of potential interactions.

4. Step 4. Time-Based Abandoned Dog Classification.

Building on the timestamped interaction data from Step 3, we apply a time-based strategy to categorize potential abandonment cases. A configurable observation threshold, denoted as T, defines how long a detected dog is tracked to determine whether a nearby person (likely its owner) remains within proximity. During this "T minutes" period, if a person consistently appears near the dog, the dog is classified as "Not Abandoned." If, however, no person remains within the proximity threshold throughout the observation period, the system flags the dog as "Abandoned." This approach enables a more reliable classification by filtering out false positives due to short-term separations or occlusions. In the experiments explained in Subsection F, we tested three different durations for T: 30, 60, and 120 seconds. Ultimately, a T value of 60 seconds was selected, as it offered the best trade-off between efficiency and detection, minimizing misclassifications without compromising early detection.

5. Step 5: Classifying and Responding to Abandoned Dogs.

If a dog is not accompanied by a potential owner after the "T minutes" observation period, it is classified as "Abandoned".

2. Pseudo-Code for the Abandoned Dog Classification Workflow

We propose the following pseudo-code, as it can be seen in Algorithm 1, for the abandoned dog classification workflow to effectively detect and classify abandoned dogs in real-time. This algorithm outlines the complete process designed to detect and classify potentially abandoned dogs in real-time. The details of each step are explained in the previous subsection.

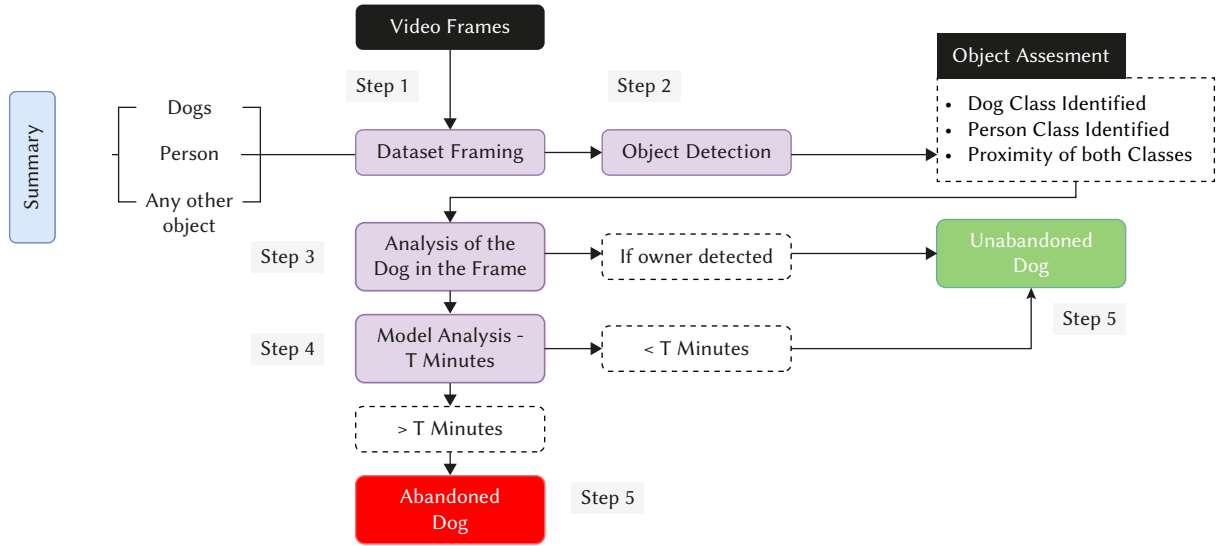


Fig. 5. System workflow for detecting abandoned dogs in natural spaces.

E. Machine Learning and Deep Learning Models for Abandoned Dog Classification

This study evaluates four machine learning and deep learning models for abandoned dog classification. These models include the YOLO models (YOLOv5, YOLOv6, and YOLOv8) and Faster R-CNN. Each model relies on a backbone network architecture that serves as a feature extractor, allowing the system to analyze input images and identify relevant patterns for object detection.

Table I summarizes the major components of these algorithms, including their backbone designs and the 32: deep learning frameworks used for implementation. This comparative overview helps to contextualize their performance in terms of classification performance, speed, and suitability for real-time detection tasks.

TABLE I. OBJECT IDENTIFICATION ALGORITHMS, THEIR BACKBONES, AND FRAMEWORKS

Algorithm	Backbone	Framework
YOLOv5	CSPDarknet53	Ultralytics (PyTorch)
YOLOv6	EfficientNet-L2 + SPP	Meituan (PyTorch)
YOLOv8	CSPDarknet + self-attention, improved FPN	Ultralytics (PyTorch)
Faster R-CNN	ResNet-50 + FPN	Facebook's Detectron2 (PyTorch)

Starting from the YOLO models, YOLOv5 is based on the CSPDarknet53 architecture, offers multiple variants, and is implemented using the PyTorch framework. YOLOv6 developed by Meituan incorporates EfficientNet-L2 and Spatial Pyramid Pooling (SPP) to enhance feature extraction. YOLOv8, developed by Ultralytics, builds on previous versions by integrating self-attention mechanisms and an improved feature pyramid network into the CSPDarknet architecture.

Faster R-CNN employs ResNet-50 as its backbone and incorporates a Region Proposal Network (RPN) to enhance object detection classification performance. In this study, we implemented Faster R-CNN using Facebook's Detectron2, a PyTorch-based framework that integrates ResNet with a Feature Pyramid Network (FPN), along with standard RPN and Region of Interest (ROI) heads, to enable precise and scalable object detection.

1. Object Detection Using the YOLO Models

Here we present the YOLO models employed in this study (YOLOv5, YOLOv6, and YOLOv8). The YOLO approach mirrors aspects of the human visual system, leveraging contextual cues to quickly identify and localize relevant objects within a scene.

- 1. Labelling Strategy:** To properly identify a dog as abandoned or monitored, we developed a labelling technique that includes contextual information. YOLO models identify objects using a grid-based technique. Each grid cell is responsible for predicting the presence of an object, assigning confidence scores, and recommending bounding boxes to frame the detected objects. To replicate the human brain's capacity to distinguish objects and their environment, we used a focused labelling method.

(a) **Abandoned Dog Class:** We used loose bounding boxes around dogs suspected of abandonment. This strategy captures not just the dog but also its immediate background, allowing the model to learn from the environmental context and provide better cues about abandonment. The rationale aligns with how humans make inferences based on surrounding cues rather than focusing solely on the object.

(b) **Monitored Dog Class:** For monitored dogs, which are accompanied by a human, we used tight bounding boxes, focusing on precision. In this case, we considered the dog and the nearby human as a single compound object to help the model learn association patterns between the two. This technique ensures that the network can distinguish between isolated dogs and those under human supervision.

- 2. YOLO Models Details:** We started with YOLOv5. Its compact design enables quicker inference while maintaining classification performance. We used YOLOv6 and YOLOv8 in our research to handle larger datasets and generate more accurate bounding boxes with greater localization. This made them very efficient for differentiating nuanced situations of abandonment.

To guarantee accurate classification and localization, models were trained using the following parameters:

- Image size for YOLOv5 is 416x416 pixels, whereas YOLOv6 and YOLOv8 are 640x640 pixels each.
- Batch size is 16 for YOLOv5 and 10 for YOLOv6/v8.
- Epochs are 200 for YOLOv5/v6 and 100 for YOLOv8.

(d) **Regularization:** We used weight decay of 0.0005 to prevent overfitting and dropout methods (at 0.5) to increase generalization.

3. **Managing Partial Detection and Background Confusion with Context-Aware Bounding Boxes:** The combination of loose and tight bounding boxes allowed the model to collect both direct and contextual information in the environment, which is critical for distinguishing between abandoned and monitored dogs. In situations when grid cells partially overlap between the object and the backdrop, YOLO models are prone to error. To address this issue, we improved the ground truth bounding boxes to include elements of the background, allowing the model to integrate background information without misclassifying objects. This improvement is especially effective for finding abandoned dogs in contexts where the distinction between foreground (dog) and background (isolated places) is critical.

2. Object Detection Using Faster R-CNN for Abandoned Dog Monitoring

The Faster R-CNN is a two-stage architecture for object identification. First, it creates region proposals, which are object-bounding boxes containing objects. Second, it uses a CNN to classify and improve the location of these recommendations.

1. **Region Proposal Generation:** The R-CNN architecture starts by finding potential areas within an image that might contain objects. For this work, we employed a selective search method that produces around 2000 region ideas for every image. These recommendations serve as prospective bounding boxes for concepts such as "abandoned dogs" and "monitored dogs." Selective search was chosen because it provides an appropriate mix of speed and proposal quality. It blends superpixels based on texture, colour, and size similarities, allowing the model to capture regions of interest without the need for previous annotations during the proposal generation step. While slower than the more RPNs employed in Faster R-CNN, selective search has a high recall rate, which is critical for recognising tiny objects like dogs in crowded or complicated surroundings.
2. **Feature Extraction and Classification:** Once the region proposals develop, each candidate region is fed into a pre-trained CNN, ResNet50, to extract deep features that characterize its visual properties. ResNet50 was chosen due to its deep residual architecture, which captures complicated patterns necessary for identifying between similar classes, such as "abandoned dogs" and "monitored dogs." The extracted characteristics are then categorized using a SVM (Support Vector Machine) classifier. Each idea falls under one of three categories: "abandoned dog," "monitored dog," or "background." Following classification, a bounding box regression step is used to adjust the predicted bounding box coordinates, ensuring the detected object's precise localization. Transfer learning was used to fine-tune the later layers of ResNet50 after freezing the first layers. This method ensures that the model can detect not just dogs but also the surrounding context, which may indicate abandonment.
3. **Labelling Strategy:** Each image was marked with bounding boxes labelled "abandoned dog" or "monitored dog." For the abandoned dog class, loose bounding boundaries have been established to incorporate both the dog and its surrounding surroundings, allowing the model to learn from visual clues that may suggest abandonment. To distinguish between abandoned dogs and those under supervision, tighter bounding boxes were used to contain the dog and its human companion, which were considered as a single object. Data preparation involved the following steps:

(a) **Annotation:** Bounding boxes were labelled "abandoned dog" or "monitored dog," ensuring that all objects were covered and the surrounding context was taken into account for improved model learning.

(b) **Normalization:** Images were scaled to a consistent 600x600 pixel format to standardize the CNN's input dimensions.

4. **Hyperparameters and Training:** To improve model performance, numerous hyperparameters were modified. To avoid overfitting and maintain stable convergence, the learning rate was set at 0.001, with a 0.1 decay factor added after every 10 epochs. The batch size of 8 was chosen to strike a compromise between computational efficiency and memory restrictions. Weight decay of 0.0005 and a dropout rate of 0.5 were used to regularize the model, which was especially important given the dataset's small size. The statistics showed a class imbalance, with more cases of monitored dogs than abandoned dogs. To counter this, strategies such as oversampling the minority class and class-specific weighting during training were used. These tactics guaranteed that the model remained responsive to uncommon cases of abandoned pets while not overfitting to the majority class.

F. Experiments

To evaluate and test our methodology, we conducted a series of experiments. All experiments were performed in a standardized computational environment using Google Colab Pro, which provides access to an NVIDIA Tesla T4 GPU.

All the experiments followed the classification process based on the steps described in Algorithm 1. The machine learning and deep learning models used, along with their configurations, are those described in Subsection E. Below, we provide a description of the experiments.

1. Experiment 1: Dog-Human Interaction Analysis for Abandonment Detection Using YOLO Models

In this experiment, we evaluated the performance of three YOLO models (YOLOv5, YOLOv6, and YOLOv8) in detecting dogs and humans in video frames, with a specific focus on analyzing their co-occurrence as an indicator of potential abandonment. The objective was to determine how accurately each model could identify the presence of dogs, and dogs and humans within the same frame and support the inference of abandonment based on the absence of interaction. This experiment provides insights into the strengths and limitations of each model in detecting abandonment scenarios under natural conditions.

2. Experiment 2: Impacts of Single-Channel Color Image Augmentation on Model Performance

We examined the impact of image augmentation on model performance by exploring a single-color channel augmentation technique across datasets in image recognition tasks. The experiment focused on applying this technique to a small collection of photos depicting abandoned dogs in natural settings before and after applying single-color channel image augmentation. The experiment was conducted using YOLOv5, which was selected for its reliability and consistency, providing a stable baseline to isolate and assess the effects of the augmentation strategy. Below, we describe how the datasets of the small collection of photos used in this experiment were constructed.

- **Dataset:** We started with a collection of seven distinct batches of 500 images of abandoned dogs in natural surroundings, referred to as Samples ranging from Sample 1 to Sample 7. These images included various breeds and sizes of dogs, captured under different lighting conditions and environmental settings.
- **Data Splitting:** Each Sample was split into 70% training, 20% testing

sets and 10% validation sets, resulting in 350 training images, 100 testing images and 50 validation images.

3. Experiment 3: Performance Evaluation of Our Workflow Abandoned Dog Detection

In this experiment, we evaluated the performance of the workflow presented in Subsection D. We evaluated the performance employing the YOLO models (YOLOv5, YOLOv6, and YOLOv8) and the Faster R-CNN model. We used transfer learning on the custom dataset to take advantage of pre-trained weights and simplify the training process. This experiment also served to test if single-stage detection models like the YOLOs outperform multi-stage approaches such as Faster R-CNN in terms of inference speed while maintaining comparable classification performance.

4. Experiment 4: Ablation Study on the Abandoned Dog Detection System

To better understand the individual contributions of the main components within our abandoned dog detection system, we designed an ablation study based on the systematic exclusion of key elements. This experiment assesses the relative importance of each component in the overall performance of the system, thus validating their inclusion and role in the final workflow. We employed YOLOv5, YOLOv6, and YOLOv8 models as the baseline detection architectures.

The components evaluated in the ablation study are:

- *Trajectory Analysis Using Optical Flow and Dwell-Time Thresholds*: By isolating this component, we aim to determine the extent to which temporal behavioral cues contribute to accurate classification and the reduction of false positives.
- *Single-Channel Image Augmentation in HSV Color Spaces*: The ablation of this component allows us to evaluate its effect on robustness across diverse visual conditions.

G. Performance Evaluation Metrics

In this subsection we present the metrics employed to assess the real-world applicability of the YOLO models and Faster R-CNN for abandoned dog detection.

1. Detection Metrics for YOLO Models

- **Precision**: This metric reflects how many of the dogs identified as abandoned are truly abandoned. High precision ensures our model minimizes false positives, which are non-abandoned dogs mistakenly classified as abandoned.
- **Recall**: This metric addresses how well our model captures all the actual abandoned dogs in the scene. High recall minimizes false negatives, which are actual abandoned dogs our model misses.
- **Mean Average Precision (mAP)**: This metric combines precision and recall, providing an overall measure of our model's detection accuracy. A higher mAP indicates our image augmentation techniques are successful in improving abandoned dog detection. In this paper, we report mAP in conjunction with **Intersection over Union (IoU)**. IoU measures the overlap between the predicted bounding box and the ground truth bounding box, and it is calculated as the area of overlap divided by the area of union. A higher IoU score indicates a more accurate localization. mAP was computed using two thresholds: **mAP@0.5**, which uses an IoU threshold of 0.5, and **mAP@0.5:0.95**, which averages mAP over IoU thresholds ranging from 0.5 to 0.95.
- **Confidence Scores**: The confidence scores relate to the "loose" bounding boxes. These scores reflect the model's certainty about classifying a dog with surrounding context as abandoned. Higher confidence scores suggest the dog is more likely to be abandoned,

considering the surrounding visual cues captured in the "loose" bounding box.

- **Inference Time**: This metric refers to the time in seconds required by a model to analyze an input frame and generate detection outputs, such as identifying objects like dogs and humans in an image. It is a key metric for assessing the model's suitability for real-time detection scenarios.

2. Detection Metrics for Faster R-CNN

- **fast_rcnn/cls_accuracy**: classification accuracy of the Fast R-CNN model at each iteration. It shows how accurately our model is classifying the dogs and persons into the correct categories. High classification accuracy indicates our model's proficiency in distinguishing between different objects within the scene.
- **fast_rcnn/fg_cls_accuracy**: classification accuracy on the foreground (object) samples. It shows how well the model is performing on actual objects (dogs and persons), indicating the model's effectiveness in correctly identifying the objects of interest.
- **loss_cls: classification efficiency**. This is a measure of how well the model is classifying objects into the correct categories (dogs and persons). Lower classification loss signifies better model performance in distinguishing between different objects.
- **loss_box_reg**: loss associated with bounding box regression. This measures how accurately the model is predicting the bounding boxes around detected objects (dogs and persons). Lower bounding box regression loss indicates better precision in object localization.
- **total_loss**: total loss combining all the individual losses (classification loss, bounding box regression loss, RPN classification loss, and RPN localization loss). This provides an overall measure of the model's performance during training, indicating the effectiveness of the training process in improving model accuracy.
- **rpn/num_pos_anchors**: number of positive anchors (anchors that contain objects) used by the RPN. It indicates the number of regions proposed by the RPN that contain objects (dogs and persons), essential for accurate object detection.
- **roi_head/num_bg**: The number of background (non-object) samples used by the Region of Interest (RoI) head during training. A high number of background samples indicates a large portion of the data consists of non-object regions, ensuring the model learns to differentiate between objects (dogs and persons) and non-objects effectively.
- **roi_head/num_fg**: number of foreground (object) samples used by the RoI head during training. This metric indicates how many actual object samples (dogs and persons) are being considered, ensuring the model learns to identify the objects of interest accurately.
- **rpn/num_neg_anchors**: number of negative anchors (anchors that do not contain objects) used by the RPN. This helps in balancing the training by including non-object regions, aiding the RPN in proposing regions that are more likely to contain objects of interest.
- **data_time**: This metric measures the average time (in seconds) taken to load a batch of data during training. It's crucial for identifying potential bottlenecks in the data pipeline. High data_time values may indicate that data loading is slowing down the training process, suggesting a need for optimization in data preprocessing or loading mechanisms.
- **eta_seconds**: This metric estimates the remaining time (in seconds) for the training process to complete. It's calculated by multiplying the median iteration time over a recent window.

- **fast_rcnn/false_negative**: This metric counts the number of instances where the model failed to detect an object that is present in the ground truth. A high value indicates that the model is missing many true objects, which can be critical in applications like abandoned dog detection.
- **loss_rpn_cls**: This loss measures how well the RPN distinguishes between object (foreground) and non-object (background) regions. It's calculated using binary cross-entropy loss over all anchors labeled as foreground or background. A lower value indicates better performance in proposing relevant regions for further processing.
- **loss_rpn_loc**: This loss evaluates the accuracy of the RPN in predicting the bounding box coordinates of proposed regions. It's computed using the L1 loss between the predicted and ground truth box coordinates for positive anchors. A lower value signifies more precise localization of proposed regions.
- **lr**: This metric records the average time taken to complete one training iteration, including data loading, forward pass, backward pass, and optimization steps. It's useful for assessing the training efficiency and identifying potential bottlenecks in the pipeline.
- **time**: The average time, in seconds, taken to complete a single training iteration (forward and backward pass) during model training.
- **Inference Time**: This metric refers to the time in seconds required by a model to analyze an input frame and generate detection outputs.

IV. RESULTS

In this section, we show the results of the four experiments performed.

A. Experiment 1: Dog-Human Interaction Analysis for Abandonment Detection Using YOLO Models

In this experiment we evaluated the performance of YOLO models for identifying dogs and humans in video frames by analysing their interactions to identify likely abandonment scenarios.

Table II presents a performance comparison of YOLOv5, YOLOv6, and YOLOv8 in detecting dogs and identifying interaction presence, where a dog and a human are simultaneously detected. For dog detection, YOLOv8 achieves the best overall performance, with the highest precision (0.83), recall (0.87), and mAP@0.5 and mAP@0.5:0.95 score 0.89 and 0.51 respectively. YOLOv5 follows closely behind, while YOLOv6 shows slightly lower scores across all metrics. In person detection, YOLOv8 again outperforms the others, with the highest precision (0.84), recall (0.88), and mAP@0.5 and mAP@0.5:0.95 values (0.90 and 0.49 respectively). YOLOv6 performs better than YOLOv5 in this task, particularly in recall and mAP. Regarding inference speed, YOLOv8 is the fastest, with an inference time of 0.015 seconds per frame, followed by YOLOv6 (0.019) and YOLOv5 (0.022), highlighting YOLOv8's efficiency for real-time applications.

For person-dog interaction detection, YOLOv8 again achieves the best overall performance. It records the highest precision (0.84), recall (0.88), and mAP scores (mAP@0.5 of 0.90 and mAP@0.5:0.95 of 0.49), indicating a superior ability to detect frames containing both a person and a dog. YOLOv6 follows closely, with slightly lower but still strong performance across all metrics: precision (0.81), recall (0.85), mAP@0.5 (0.86), and mAP@0.5:0.95 (0.46). YOLOv5, while still effective, shows the lowest values among the three models for interaction detection, with a precision of 0.79, recall of 0.83, mAP@0.5 of 0.85, and mAP@0.5:0.95 of 0.44.

TABLE II. PERFORMANCE COMPARISON OF YOLO MODELS ON DETECTING DOGS AND HUMANS ROUNDED TO 2 DECIMALS

Metric	YOLOv5	YOLOv6	YOLOv8
Dog Detection			
Precision	0.81	0.78	0.83
Recall	0.85	0.82	0.87
mAP@0.5	0.86	0.84	0.89
mAP@0.5:0.95	0.47	0.45	0.51
Person-Dog Interaction Detection			
Precision	0.79	0.81	0.84
Recall	0.83	0.85	0.88
mAP@0.5	0.85	0.86	0.90
mAP@0.5:0.95	0.44	0.46	0.49
Confidence Score	0.25	0.25	0.25

These results suggest that while all models are capable of identifying dog-human interaction, YOLOv8 offers the best trade-off between classification performance and efficiency.

B. Experiment 2: Impacts of Single-Channel Color Image Augmentation

Here, we present the results for experiment 2 measuring the impact of image augmentation on YOLOv5 model performance.

Table III presents a performance comparison of YOLOv5 on the seven batches of the dataset (from Sample 1 to Sample 7) before and after applying single-color channel image augmentation. Precision and recall are reported for the initial and the augmented dataset, allowing distinct evaluation perspectives on model behavior.

TABLE III. PRECISION AND RECALL COMPARISON FOR INITIAL AND AUGMENTED DATASETS USING YOLOv5 ROUNDED TO 2 DECIMALS. (PRECISION INITIAL DATASET: PREC. INIT.; PRECISION AUGMENTED DATASET: PREC. AUG.; RECALL INITIAL DATASET: REC. INIT.; RECALL AUGMENTED DATASET: REC. AUG.)

Sample	Prec. Init.	Prec. Aug.	Rec. Init.	Rec. Aug.
1	0.83	0.94	0.87	0.78
2	0.73	0.55	0.87	0.94
3	0.96	0.83	0.77	0.91
4	1.00	0.99	0.83	0.87
5	0.63	0.94	0.79	0.79
6	0.69	0.91	0.88	0.76
7	1.00	0.96	0.57	0.80

The Sample sets showed notable precision gains after augmentation. For instance, Sample 1 improved from 0.83 to 0.94, and Sample 5 from 0.63 to 0.94, indicating enhanced detection of true positives. Even Sample 6 exhibited a strong gain, rising from 0.69 to 0.91.

Instead, Samples exhibited more variable results in terms of recall. While Sample 2 and Sample 3 saw substantial improvements, from 0.87 to 0.94 and from 0.77 to 0.91, respectively, others like Sample 1 and Sample 6 experienced slight drops (from 0.87 to 0.78 and from 0.88 to 0.76).

Additionally, Fig. 6 shows a grouped bar chart comparing the performance metrics of the model trained on the initial and the augmented dataset. The chart includes four bars labeled "Precision Initial Dataset," "Precision Augmented Dataset," "Recall Performance on Initial Dataset," and "Recall Performance on Augmented Dataset." Each bar is divided into two sections: a darker segment representing the weighted average and a lighter segment representing the macro average. The precision values for the initial dataset are both 0.83, while for the augmented dataset they are 0.88 (weighted) and 0.87 (macro). The recall values are 0.80 for both averages in the initial dataset and 0.84 (weighted) and 0.83 (macro) for the augmented dataset.

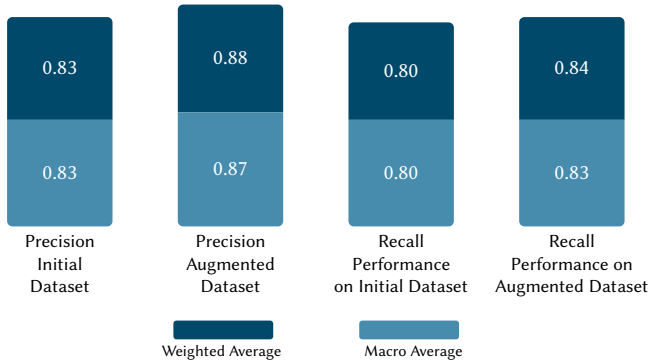


Fig. 6. Graphical Comparison of Precision and Recall for Initial and Augmented Sets.

C. Experiment 3: Performance Evaluation of Our Workflow Abandoned Dog Detection

In this subsection, we present the results of the performance of the YOLO and the Faster R-CNN models.

1. YOLO Models Performance

Here, we present the performance of the YOLO models (YOLOv5, YOLOv6, and YOLOv8) in locating abandoned pets.

Table IV provides a comprehensive summary of the performance.

TABLE IV. SUMMARY OF YOLO MODEL PERFORMANCE ANALYSIS IN ABANDONED DOGS DETECTION ROUNDED TO 2 DECIMALS

Model	YOLOv5	YOLOv6	YOLOv8
Epochs	195	200	100
Training Duration (hours)	0.30	0.12	1.77
Layers	213	142	168
Parameters	7,015,519	4,233,942	3,006,038
GFLOPs	15.80	11.80	8.10
Overall Precision	0.81	1.00	0.87
Overall Recall	0.85	0.08	0.46
mAP@0.5	0.86	0.54	0.67
mAP@0.5:0.95	0.47	0.22	0.45
Abandoned Precision	0.83	1.00	0.91
Abandoned Recall	0.87	0.07	0.67
Abandoned mAP@0.5	0.86	0.53	0.82
Abandoned mAP@0.5:0.95	0.51	0.21	0.60
Not Abandoned Precision	0.80	1.00	0.83
Not Abandoned Recall	0.84	0.11	0.26
Not Abandoned mAP@0.5	0.85	0.55	0.53
Not Abandoned mAP@0.5:0.95	0.43	0.23	0.29
Inference Time (ms/frame)	1.7	0.3	0.3

YOLOv5 trained over 195 epochs with 7 million parameters and 15.8 GFLOPs. While YOLOv5 had promising results: an overall Precision of 0.81 and Recall of 0.85, with a solid mAP@0.5 of 0.86, its performance declined due to a small dataset and difficulty with complex situations.

To solve these problems, we moved to YOLOv6. This model used a custom dataset and a more complex architecture, resulting in considerable gains. YOLOv6 was trained over 200 epochs and had fewer parameters (4.2 million) and GFLOPs (11.8). Notably, YOLOv6 attained perfect Overall Precision (1.00), resulting in an Abandoned and Not Abandoned Precision of 1.00, demonstrating its capacity to recognize all positive instances. However, a trade-off emerged: YOLOv6 had low Overall Recall (0.08), meaning that it missed a large proportion of actual abandoned dogs scenarios. This resulted in a lower mAP@0.5 (0.54%) than YOLOv5.

YOLOv8, with the lowest number of parameters (3 million) and GFLOPs (8.1), suggested a compromise between efficiency and overall detection performance. It obtained a reasonable mAP@0.5 (0.67) achieving an Abandoned mAP@0.5 of 0.82. However, the efficacy of the abandoned dog detecting system is an important concern. While YOLOv8 maintained excellent Abandoned Precision (0.91%), Abandoned Recall (0.67%) was modest, indicating that it still missed a large percentage of abandoned dogs scenarios.

In terms of inference time, YOLOv8 and YOLOv6 were equally the fastest (0.3 ms/frame), both outperforming YOLOv5 (1.7 ms/frame), which confirms their advantage in speed and real-time performance.

2. Faster R-CNN Performance

We report the performance of Faster R-CNN.

Table V provides a detailed summary of the performance of the Faster R-CNN model during training for abandoned dog detection. The table includes both cumulative (SUM) and AVERAGE values across training iterations for a variety of performance metrics.

TABLE V. SUMMARY OF FASTER R-CNN PERFORMANCE ANALYSIS IN ABANDONED DOGS DETECTION ROUNDED TO 2 DECIMALS

METRIC	SUM	AVERAGE
fast_rcnn/cls_accuracy	96.67	0.87
roi_head/num_bg	12325.00	111.03
roi_head/num_fg	1882.00	16.95
rpn/num_neg_anchors	27984.25	252.11
rpn/num_pos_anchors	431.75	3.89
data_time	8.85	0.08
eta_seconds	14595.45	132.68
fast_rcnn/false_negative	67.33	0.61
fast_rcnn/fg_cls_accuracy	20.81	0.19
loss_box_reg	44.17	0.40
loss_cls	42.31	0.38
loss_rpn_cls	0.24	0.00
loss_rpn_loc	0.56	0.00
lr	0.24	0.00
time	44.23	0.40
total_loss	88.32	0.80
Inference Time (s -ms/frame)	10.67	220

The model achieved a high average classification accuracy (fast_rcnn/cls_accuracy) of 0.87, with a total value of 96.67, indicating overall strong classification performance. The number of background samples (roi_head/num_bg) was consistently higher than the number of foreground samples (roi_head/num_fg), with averages of 111.03 and 16.95 respectively—this reflects the typical imbalance in object detection tasks.

Anchor analysis shows a similar pattern, with the number of negative anchors (rpn/num_neg_anchors) being much greater than the number of positive anchors (rpn/num_pos_anchors), averaging 252.11 and 3.89 respectively. This imbalance is common and helps the model learn to distinguish object versus non-object regions more effectively.

Regarding loss metrics, the model maintained relatively low values across classification loss (loss_cls = 0.38), box regression loss (loss_box_reg = 0.40), and RPN-related losses (loss_rpn_cls and loss_rpn_loc = 0.00). The total loss (total_loss) averaged 0.80, reflecting stable convergence during training.

Other indicators such as data loading time (data_time = 0.08) and iteration time (time = 0.40) remained consistent, suggesting good computational performance. The learning rate (lr) gradually increased during training.

In terms of Inference Time, the Faster R-CNN model achieved a total inference time of 10.67 seconds, corresponding to an average of 220 ms.

D. Experiment 4: Ablation Study on the Abandoned Dog Detection System

The results of the ablation study are presented below, starting with the effect of the trajectory analysis module, followed by the impact of the single-channel image augmentation technique.

1. Effect of Trajectory Analysis

Here we present the results after removing the trajectory analysis module.

Table VI summarizes the performance of the YOLOv5, YOLOv6, and YOLOv8 models when the trajectory analysis component was removed from the abandoned dog detection workflow.

TABLE VI. SUMMARY OF YOLO MODEL PERFORMANCE WITHOUT TRAJECTORY ANALYSIS IN ABANDONED DOGS DETECTION ROUNDED TO 2 DECIMALS

Model	YOLOv5	YOLOv6	YOLOv8
Epochs	195	200	100
Training Duration (hours)	0.30	0.12	1.77
Layers	213	142	168
Parameters	7,015,519	4,233,942	3,006,038
GFLOPs	15.80	11.80	8.10
Overall Precision	0.76	0.94	0.82
Overall Recall	0.85	0.08	0.46
mAP@0.5	0.86	0.54	0.67
mAP@0.5:0.95	0.47	0.22	0.45
Abandoned Precision	0.78	0.94	0.86
Abandoned Recall	0.87	0.07	0.67
Abandoned mAP@0.5	0.86	0.53	0.82
Abandoned mAP@0.5:0.95	0.51	0.21	0.60
Not Abandoned Precision	0.75	0.94	0.78
Not Abandoned Recall	0.84	0.11	0.26
Not Abandoned mAP@0.5	0.85	0.55	0.53
Not Abandoned mAP@0.5:0.95	0.43	0.23	0.29

YOLOv5 maintained solid recall (0.85) and balanced precision (0.76), with high mAP@0.5 (0.86), suggesting it remained fairly robust even without trajectory tracking. However, its performance on distinguishing abandoned from non-abandoned cases showed a slight decrease, particularly in precision.

YOLOv6, while achieving the highest precision (0.94 for both classes), showed extremely low recall (0.08 overall and 0.07 for abandoned cases), indicating that it missed a large number of actual abandoned dog instances. This makes the model unreliable in this context, despite its perfect predictions when confident.

YOLOv8 offered a middle ground, with a balanced performance: 0.82 precision, 0.46 recall, and 0.67 mAP@0.5. For the abandoned category, it achieved 0.86 precision and 0.67 recall, indicating good detection performance even without trajectory input.

All models experienced declines in precision, increasing the false positives, particularly for the "Not Abandoned" class.

2. Effect of Single-Channel Image Augmentation

We present the results assessing the impact of removing the data augmentation module from the system.

Table VII presents the performance results of YOLOv5, YOLOv6, and YOLOv8.

TABLE VII. YOLO MODELS PERFORMANCE WITHOUT HSV-BASED AUGMENTATION ROUNDED TO 2 DECIMALS

Model	YOLOv5	YOLOv6	YOLOv8
Epochs	195	200	100
Training Duration (hours)	0.30	0.12	1.77
Layers	213	142	168
Parameters	7,015,519	4,233,942	3,006,038
GFLOPs	15.80	11.80	8.10
Overall Precision	0.78	0.96	0.84
Overall Recall	0.82	0.08	0.44
mAP@0.5	0.83	0.52	0.64
mAP@0.5:0.95	0.45	0.21	0.43
Abandoned Precision	0.80	0.96	0.87
Abandoned Recall	0.84	0.07	0.64
Abandoned mAP@0.5	0.83	0.51	0.79
Abandoned mAP@0.5:0.95	0.49	0.20	0.58
Not Abandoned Precision	0.77	0.96	0.80
Not Abandoned Recall	0.81	0.11	0.25
Not Abandoned mAP@0.5	0.82	0.53	0.51
Not Abandoned mAP@0.5:0.95	0.41	0.22	0.28

YOLOv5 maintained a balanced performance with 0.78 precision, 0.82 recall, and 0.83 mAP@0.5, showing it remains relatively robust even without augmentation. However, slight reductions compared to the baseline (with augmentation) suggest a positive impact of HSV manipulation on its predictive quality.

YOLOv6 again shows very high precision (0.96) but extremely low recall (0.08 overall and 0.07 for abandoned cases). This confirms its tendency to predict very selectively when confidence is high, missing many actual cases. Its mAP@0.5 is also low (0.52), indicating weak generalization without augmentation.

YOLOv8 strikes a moderate balance, reaching 0.84 precision and 0.44 recall, with an abandoned mAP@0.5 of 0.79. Compared to the results with HSV-based augmentation, its performance slightly declines, showing that augmentation plays a meaningful role in boosting recall and localization consistency.

Across all models, removing HSV-based augmentation leads to noticeable drops in recall and mAP scores, particularly for YOLOv8 and YOLOv5. This confirms that color-based augmentation helps the models generalize better to varied lighting and appearance conditions.

V. DISCUSSION

A. Principal Findings

This research paper examined using computer vision and deep learning in finding abandoned dogs, with a focus on object recognition methods and data augmentation approaches. Here, we discuss how these findings answer our research questions and align with existing literature.

Our study demonstrated that deep learning-based computer vision algorithms could effectively distinguish between dogs with and without human companions in diverse real-world settings. The integration of trajectory analysis (optical flow and stay time) with image segmentation proved efficient in differentiating abandoned dogs from those with human companions based on their movement patterns and interactions with the environment. This approach facilitated accurate classification regardless of the dogs' varying appearances and surroundings, thus validating the research question: "can deep learning-based computer vision algorithms accurately distinguish between dogs with and without human companions in real-world settings (natural and urban), despite differences in appearance and surroundings?"

The implementation of single-channel image augmentation using the HSV color space significantly improved model robustness to variations in lighting and positioning, which is particularly advantageous for smaller datasets. This method enhanced the model's generalization to real-world illumination conditions, leading to improved classification performance. Our results showed a 4% increase in Recall and Precision with single-channel augmentation over default RGB or HSV images, confirming the effectiveness of data augmentation techniques and addressing the research question: "Do data augmentation approaches, including single-channel image augmentation, improve the performance of deep learning models for detecting abandoned dogs?"

The results demonstrated that YOLO models, particularly YOLOv8, are highly effective in the real-time detection of abandoned dogs when leveraging loose bounding boxes. These bounding boxes, which intentionally include surrounding contextual information, allow the model to make more informed predictions regarding abandonment scenarios. Among the tested models, YOLOv8 consistently outperforms YOLOv5 and YOLOv6 in both dog detection and person-dog interaction identification, achieving the highest precision, recall, and mean average precision values. This suggests that YOLOv8 is especially well-suited for applications that require not only object localization but also the interpretation of spatial relationships between entities, a critical factor when detecting abandonment based on proximity and co-occurrence patterns. Furthermore, the inference time results show that YOLOv8 offers superior processing speed (0.3 ms/frame), validating its potential for deployment in real-time systems. This speed, combined with high classification performance, confirms the model's capability to operate effectively under real-world constraints, where rapid response and continuous video stream analysis are essential. The performance of the models using loose bounding boxes also highlights an important trade-off: while larger bounding boxes may reduce precision in some object detection tasks, in this context, they help capture interaction zones and environmental indicators that are vital for distinguishing between abandoned and accompanied dogs. Thus, YOLO models, especially YOLOv8, prove to be both efficient and context-aware in detecting potentially abandoned dogs in dynamic environments answering the research question: "How effectively do alternative object detection algorithms (YOLO models) perform in real-time detection of abandoned dogs using 'loose' bounding boxes that incorporate contextual information?"

Additionally, we benchmarked several object detection models, including YOLO models (YOLOv5, YOLOv6, YOLOv8), and Faster R-CNN on the same dataset to reveal the advantages and disadvantages of each strategy. YOLOv8 achieved the best trade-off between efficiency and overall detection performance. It achieved a reasonable overall mean Average Precision of 0.67 at an Intersection over Union threshold of 0.5, and reached 0.82 for detecting abandoned dogs. This predicts quicker inference speeds 0.3 ms/frame than YOLOv5 and YOLOv6. However, the efficacy of the abandoned dog detecting system is an important concern. While YOLOv8 maintained excellent Precision (0.91) for detecting abandoned dogs, Recall (0.67) was modest, indicating that it still missed a large percentage of abandoned dogs scenarios. While Faster R-CNN demonstrated slightly higher precision in some static scenes, its computational complexity and higher inference time (220 ms/frame) made it less suitable for real-time deployment in natural environments.

These findings revealed that single-stage detectors like YOLOv8 achieved a superior balance of abandoned dog detection performance and speed for real-time applications. These detectors were particularly effective when using 'loose' bounding boxes that incorporated contextual information, ensuring efficient detection of abandoned dogs. In contrast, multi-stage detectors such as Faster R-CNN, although

accurate, were slower and less suitable for real-time scenarios. This comparison highlights the effectiveness of YOLOv8 in real-time detection tasks, addressing the research questions: "how do existing object detection algorithms (e.g., YOLO models, and Faster R-CNN) perform in detecting and classifying abandoned dogs?", "what is the speed and classification performance of Abandoned dog detection method in real-time settings?"

The ablation study provides insights into the contribution of two core components (trajectory analysis and color-based data augmentation technique) to the performance of the proposed abandoned dog detection system. Trajectory analysis improves temporal reasoning and minimize false positives. When removed, all models, especially YOLOv6 and YOLOv8, exhibited reduced capacity to distinguish abandoned from non-abandoned cases. YOLOv8 maintained relatively strong performance but still showed reduced reliability compared to the full workflow. This confirms that the absence of trajectory-based continuity weakens the model's ability to detect actual abandonment events highlighting the added value of temporal tracking. Single-channel HSV-based image augmentation was also found to be an important contributor to detection robustness. Its removal led to a moderate but consistent decrease in Recall and mean Average Precision across all models.

B. Comparison With Prior Work

Existing methodologies in detection of abandoned dogs in natural environments [26], [39]–[41] adapted its approach from static object detection paradigms, falling short in capturing the temporal and contextual dynamics necessary for accurate identification. This research advances a comprehensive methodological shift, integrating deep learning with spatio-temporal logic to address these shortcomings.

Unlike traditional object detectors which rely on instantaneous frame-based classification [20], this study reconceptualizes abandoned dog detection as a temporally evolving event, requiring continuous observation and dwell-time verification. The proposed architecture fuses object detection with trajectory analysis, evaluating movement patterns and stationary duration of dogs across sequential video frames. An abandonment classification is only triggered when a dog remains unaccompanied for configurable temporal thresholds, validated by spatial separation from human companions. This temporally-aware formulation significantly reduces false positives caused by transient separations or occlusions, limitations often observed in conventional YOLO and Faster R-CNN models.

Conventional object detection models employ tight bounding boxes optimized for object boundaries. Our methodology introduces the use of context-extended (loose) bounding boxes, capturing the dog and its surrounding spatial context, particularly human proximity. Inspired by surveillance methodologies for abandoned luggage detection [20], this spatial expansion allows the detection model to incorporate latent cues (e.g., absence of humans), providing a cognitive-level abstraction of abandonment.

Comparing our work with past research studies, our results are consistent with those of Chatfield et al. [48], who found a 3% decline in classification precision between grayscale and RGB pictures in their experiments on the ImageNet [29] and PASCAL VOC [49] datasets. Jurio et al. [23] also evaluated the performance of image segmentation in several color space representations, such as RGB, YUV, CMYK, and HSV. In our investigation, we found that using single color space images with CNN resulted in a 4% boost in Recall and Precision over default RGB or HSV images. This lends credence to the idea that color alterations do not always destroy vital color information and can be useful for augmentation. In contrast to some fears that single color space augmentation might diminish pixel values and obfuscate

objects, we solved this issue by using image enhancing features in CNN to maintain labels and characteristics. As a result, our model was able to reliably identify, recognize, and categorize each type of animal in the datasets, validating hypothesis one.

C. Limitations and Ethical Considerations

The practical deployment of our model in real-world settings has the potential to significantly reduce the number of abandoned pets. However, various problems must be addressed to enable successful adoption. Variability in camera quality, infrastructure costs, and our method's adaptation to various environments may all have an impact on implementation. Additionally, connection with current animal rescue or monitoring systems may improve the algorithm's performance.

A notable limitation stems from the possibility of false positives, wherein dogs separated from their owners, particularly during prolonged off-leash play, may be erroneously identified as abandoned. This problem is caused by the variety in canine behavior, which can occasionally resemble signals of abandonment, such as extended lingering or autonomous mobility, even when the dog is not genuinely abandoned. The current system lies in the fixed time threshold used to determine whether a dog is considered abandoned. We chose to limit the evaluation to thresholds of 30, 60, and 120 seconds based on the duration of the available video datasets, which did not support meaningful analysis at significantly longer thresholds. This period, while essential for triggering detection events, lacks empirical grounding and may not reflect real-world human-dog interaction patterns. A static duration may lead to misclassifications, either by falsely flagging temporary separations as abandonment or by missing genuine abandonment cases due to overly conservative thresholds. Without further contextual understanding, this time parameter introduces a trade-off between prompt detection and reliability. Although our suggested methodology yields encouraging results, more efforts are required to reduce misclassification risks, enhance domain adaptation, and provide frameworks that address ethical compliance and responsible AI implementation.

We also acknowledge that our dataset is geographically constrained to Catalonia, Spain, which limits generalizability. Due to this constraint, true cross-site validation was not feasible. Expanding the dataset to include varied geographic and cultural contexts will be essential for assessing the robustness and adaptability of the proposed method on a broader scale.

VI. CONCLUSION

Our study underscores the transformative potential of computer vision and deep learning in for detecting abandoned dogs' detection. We improved classification performance and real-time processing significantly by utilizing modern data augmentation and object identification techniques. With little data, single-channel augmentation proved successful.

Our study provides a detailed description of the implemented methodology, offering a valuable resource for researchers facing similar data size limitations in computer vision tasks. This methodology leverages state-of-the-art image augmentation techniques to effectively address the limitations of the original dataset.

The results also revealed that YOLOv8 achieves a strong trade-off between classification performance and processing speed, making it ideal for real-time applications. In contrast, multi-stage detectors like Faster R-CNN despite their excellent classification performance, were less useful for real-time detection due to slower processing speeds.

Finally, we demonstrated the benefits of incorporating time-based metrics. The proposed approach is also easily adaptable to broader contexts, such as surveillance for identifying loitering individuals, detecting abandoned objects in public spaces like airports, or monitoring wildlife in environments where spatial context is critical. Future work, explored in detail in the next section, may extend this line of research to address broader challenges in temporal video analysis, particularly those involving object persistence over time.

This study lays a solid foundation for the development of real-time, adaptive systems that support both animal welfare and public safety across diverse environments.

VII. FUTURE WORK

Expanding the scope of this research to encompass a wider range of abandoned pets, such as cats, is a promising avenue for future investigation and can significantly increase the impact of the proposed computer vision system. Additionally, actively counting human presence in natural environments could provide valuable insights into factors contributing to pet abandonment, biodiversity monitoring using computer vision and inform targeted interventions. Furthermore, given the increasing prevalence of invasive species, applying computer vision technologies to their management, as exemplified by the wolf population in the Iberian Peninsula, presents a potentially impactful research direction. These extensions will contribute to a more comprehensive understanding of the challenges associated with pet abandonment and inform the development of effective interventions for both animal welfare and public safety.

Secondly, we recognize the need to broaden the geographic scope of our dataset beyond Catalonia to enable meaningful cross-site validation and improve generalization. Building upon our current framework, the proposed methodology could be extended to support the identification and monitoring of other abandoned or invasive animal species, such as cats, rabbits, or wild canines like foxes and wolves. Each of these species presents unique behavioral and physical traits, requiring tailored detection models and classification strategies. Expanding the approach in this direction would enhance its applicability in biodiversity monitoring and urban wildlife management.

Thirdly, we acknowledge that our dataset is geographically constrained to Catalonia, Spain, hence true cross-site validation was not feasible. Our proposed methodology can be used for the identification and monitoring of more abandoned or invading animal species, including cats, rabbits, and wild canines such as foxes and wolves. Each species exhibits distinct behavioural and physical characteristics, requiring customised detection models and classification criteria. Extending the model in this approach will greatly benefit broader biodiversity monitoring and urban wildlife control applications. In addition, leveraging advanced trajectory-based metrics, such as proximity patterns, dynamically computed interaction duration, movement consistency, entropy of movement, and spatio-temporal densities, could enhance the system's ability to distinguish between brief separations and actual abandonment events.

Fourthly, non-visual signals, like as auditory signatures (e.g., barking patterns, distress vocalisations) or temperature data, could enhance the classification performance of abandoned animal identification, particularly in low-light or visually complicated contexts. The combination of multimodal data with vision-based detection can considerably increase real-world dependability.

CREDIT AUTHORSHIP

Alberto Tena: Methodology, Formal analysis, Resources, Writing -original draft, Validation.

Francesc Solsona: Writing review & editing, Supervision, Resources, Project administration, Investigation, Validation, Funding acquisition.

Javier Mora: Conceptualization, Formal analysis, Data curation, Resources, Software, Investigation, Visualization, Validation.

Oluwakemi Akinwehinmi: Wrote the paper, Idea Formulation, Conceptualization, Data curation, Resources, Software, Visualization, Validation.

Pedro Arnau: Idea Formulation, Conceptualization and designed the analysis, contributed analysis tools.

DECLARATION OF CONFLICTS OF INTEREST

The research paper authors declare that they have no competing interests.

ACKNOWLEDGMENT

This work has been supported by the Generalitat de Catalunya for the financial support to the primary author, beneficiary of a pre-doctoral grant funded under the Program contract between the Administration of the Generalitat of Catalonia, through the Department of Territory and Sustainability and the Department of Business and Knowledge, and the International Center for Numerical Methods in Engineering (CIMNE), for the period 2020-2023. This research was also been developed within the PIKSEL project, "Portal for the integration of knowledge for a sustainable ecosystems and land management" funded by Generalitat de Catalunya, through the Department of Territory and Sustainability and the Department of Climate Action. The authors also acknowledge the financial support through the Severo Ochoa Centers of Excellence Program (CEX 2018-000797-S) funded by MCIN/AEI/10.13039/501100011033

REFERENCES

- [1] W. Hu, T. Tan, L. Wang, S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 3, pp. 334–352, 2004, doi: <https://doi.org/10.1109/TSMCC.2004.829274>.
- [2] J. Redmon, A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018, doi: <https://doi.org/10.48550/arXiv.1804.02767>.
- [3] J. Hughes, D. W. Macdonald, "A review of the interactions between free-roaming domestic dogs and wildlife," *Biological Conservation*, vol. 157, pp. 341–351, 2013, doi: <https://doi.org/10.1016/j.biocon.2012.07.005>.
- [4] World Health Organization, "Rabies," 2023. Accessed: May 28, 2025, [Online]. Available: <https://www.who.int/news-room/factsheets/detail/rabies>.
- [5] K. Hampson, L. Coudeville, T. Lembo, M. Sambo, A. Kieffer, M. Atllan, J. Barrat, J. D. Blanton, D. J. Briggs, S. Cleaveland, P. Costa, C. M. Freuling, E. Hiby, L. Knopf, F. Leanes, F.-X. Meslin, A. Metlin, M. E. Miranda, T. Müller, L. H. Nel, S. Recuenco, C. E. Rupprecht, C. Schumacher, L. Taylor, M. A. N. Vigilato, J. Zinsstag, J. Dushoff, "Estimating the global burden of endemic canine rabies," *PLoS Neglected Tropical Diseases*, vol. 9, no. 4, p. e0003709, 2015, doi: <https://doi.org/10.1371/journal.pntd.0003709>.
- [6] G. Jocher, "Yolov5 by ultralytics," 2020. doi: <https://doi.org/10.5281/zenodo.3908559>.
- [7] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, X. Wei, "Yolov6: A single-stage object detection framework for industrial applications," *arXiv preprint arXiv:2209.02976*, 2022, doi: <https://doi.org/10.48550/arXiv.2209.02976>.
- [8] R. Girshick, "Fast r-cnn," in *Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 2015, pp. 1440–1448.
- [9] S. Ren, K. He, R. Girshick, J. Sun, "Faster rcnn: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017, doi: <https://doi.org/10.1109/TPAMI.2016.2577031>.
- [10] R. Girshick, I. Radosavovic, G. Gkioxari, P. Dollár, K. He, "Detectron," 2018. Accessed: May 28, 2025, [Online]. Available: <https://github.com/facebookresearch/detectron>.
- [11] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, R. Girshick, "Detectron2," 2019. Accessed: May 28, 2025, [Online]. Available: <https://github.com/facebookresearch/detectron2>.
- [12] L. Ferrando, G. Gera, C. Regazzoni, "Classification of unattended and stolen objects in videosurveillance systems," in *Proceedings of the 2006 IEEE International Conference on Video and Signal Based Surveillance*, Sydney, Australia, 2006, pp. 21–21.
- [13] F. I. Bashir, A. A. Khokhar, D. Schonfeld, "Realtime motion trajectory-based indexing and retrieval of video sequences," *IEEE Transactions on Multimedia*, vol. 9, no. 1, pp. 58–65, 2007, doi: <https://doi.org/10.1109/TMM.2006.886346>.
- [14] L. Taylor, G. Nitschke, "Improving deep learning using generic data augmentation," *arXiv preprint arXiv:1708.06020*, 2017, doi: <https://doi.org/10.48550/arXiv.1708.06020>.
- [15] C. Shorten, T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 1, p. 60, 2019, doi: <https://doi.org/10.1186/s40537-019-0197-0>.
- [16] CIMNE, "PIKSEL: Portal for the Integration of Knowledge for Sustainable Ecosystems and Land Management," 2024. Accessed: May 28, 2025, [Online]. Available: <https://pikselweb.cimne.com/>.
- [17] B. Benjdira, T. Khurshed, A. Koubaa, A. Ammar, K. Ouni, "Car detection using unmanned aerial vehicles: Comparison between faster r-cnn and yolov3," *arXiv preprint arXiv:1812.10968*, 2018, doi: <https://doi.org/10.48550/arXiv.1812.10968>.
- [18] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, Q. V. Le, "Autoaugment: Learning augmentation policies from data," *arXiv preprint arXiv:1805.09501*, 2019, doi: <https://doi.org/10.48550/arXiv.1805.09501>.
- [19] S. Lim, I. Kim, T. Kim, C. Kim, S. Kim, "Fast autoaugment," in *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, Vancouver, BC, Canada, 2019.
- [20] G. Gibbins, G. Newsam, M. Brooks, "Detecting suspicious background changes in video surveillance of busy scenes," in *Proceedings of the Third IEEE Workshop on Applications of Computer Vision (WACV'96)*, Sarasota, FL, USA, 1996, pp. 22–26.
- [21] A. H. Freedman, I. Gronau, R. M. Schweizer, D. Ortega-Del Vecchyo, E. Han, P. M. Silva, et al., "Genome sequencing highlights the dynamic early history of dogs," *PLoS Genetics*, vol. 10, no. 1, p. e1004016, 2014, doi: <https://doi.org/10.1371/journal.pgen.1004016>.
- [22] M. E. Gompper, "The dog-human-wildlife interface: Assessing the scope of the problem," in *Free-Ranging Dogs and Wildlife Conservation*, Oxford University Press, 2013.
- [23] S. Barnard, C. Ippoliti, D. Di Flaviano, A. De Ruvo, S. Messori, A. Giovannini, P. Dalla Villa, "Smartphone and gps technology for freeroaming dog population surveillance - a methodological study," *Veterinaria Italiana*, vol. 51, no. 3, pp. 165–172, 2015, doi: <https://doi.org/10.12834/VetIt.233.2163.3>.
- [24] P. Howell, *At Home and Astray: The Domestic Dog in Victorian Britain*. Charlottesville, VA, USA: University of Virginia Press, 2015.
- [25] O. Akinwehinmi, P. Arnau, F. Solsona, A. Priegue, J. Jimenez, "Paws for concern: Spotting strays and abandoned pets in natural spaces," in *2023 IEEE International Smart Cities Conference (ISC2)*, Bucharest, Romania, 2023, pp. 1–7.
- [26] J. Ruiz-Chavez, J. Salvador-Meneses, C. MejíaAstudillo, S. Diaz-Quilachamin, "Analysis of dogs' abandonment problem using georeferenced multiagent systems," in *From Bioinspired Systems and Biomedical Applications to Machine Learning*, Cham, Switzerland: Springer International Publishing, 2019, pp. 297–306.
- [27] B. Graham, "Fractional max-pooling," *arXiv preprint arXiv:1412.6071*, 2015, doi: <https://doi.org/10.48550/arXiv.1412.6071>.
- [28] Ruman, "Yolo data augmentation explained: Turbocharge your object detection model," 2023. Accessed: May 28, 2025, [Online]. Available: <https://www.rumn.ai/blog/yolodata-augmentation-explained>.
- [29] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, F.F. Li, "Imagenet: A large-scale

hierarchical image database,” in *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, 2009, pp. 248–255.

[30] O. Akinwehinmi, P. Arnau, F. Solsona, “Improving abandoned pet detection in natural spaces through adaptive single-coloured image augmentation techniques,” in *Proceedings of the 15th Congress of the International Colour Association (AIC 2023)*, Chiang Rai, Thailand, 2023.

[31] T. DeVries, G. W. Taylor, “Improved regularization of convolutional neural networks with cutout,” *arXiv preprint arXiv:1708.04552*, 2017, doi: <https://doi.org/10.48550/arXiv.1708.04552>.

[32] H. Zhang, M. Cisse, Y. N. Dauphin, D. Lopez-Paz, “Mixup: Beyond empirical risk minimization,” *arXiv preprint arXiv:1710.09412*, 2018, doi: <https://doi.org/10.48550/arXiv.1710.09412>.

[33] Y. Guo, Y. Mao, R. Zhang, “Mixup as locally linear out-of-manifold regularization,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, Honolulu, HI, USA, 2019, pp. 3714–3722.

[34] S. Yun, D. Han, S. Chun, S. J. Oh, Y. Yoo, J. Choe, “Cutmix: Regularization strategy to train strong classifiers with localizable features,” in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, South Korea, 2019, pp. 6022–6031.

[35] A. Bruno, E. Ardizzone, S. Vitabile, M. Midiri, “A novel solution based on scale invariant feature transform descriptors and deep learning for the detection of suspicious regions in mammogram images,” *Journal of Medical Signals and Sensors*, vol. 10, no. 3, pp. 158–173, 2020, doi: https://doi.org/10.4103/jmss.JMSS_31_19.

[36] B. Recht, R. Roelofs, L. Schmidt, V. Shankar, “Do cifar-10 classifiers generalize to cifar10?,” *arXiv preprint arXiv:1806.00451*, 2018, doi: <https://doi.org/10.48550/arXiv.1806.00451>.

[37] A. Jurio, M. Pagola, M. Galar, C. LopezMolina, D. Paternain, A comparison study of different color spaces in clustering based image segmentation,” in *Information Processing and Management of Uncertainty in Knowledge-Based Systems. Applications*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 532–541.

[38] Y. Bengio, F. Bastien, A. Bergeron, N. Boulanger-Lewandowski, T. Breuel, Y. Chherawala, M. Cisse, M. Côté, D. Erhan, J. Eustache, X. Glorot, X. Muller, S. Pannetier Lebeuf, R. Pascanu, S. Rifai, F. Savard, G. Sicard, “Deep learners benefit more from out-of-distribution examples,” in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 15 of Proceedings of Machine Learning Research, Fort Lauderdale, FL, USA, 2011, pp. 164–172.

[39] E. Azizi, L. Zaman, “Deep learning pet identification using face and body,” *Information*, vol. 14, no. 5, 2023, doi: <https://doi.org/10.3390/info14050278>.

[40] P. Kasnesis, V. Doulerakis, D. Uzunidis, D. G. Kogias, S. I. Funcia, M. B. González, C. Giannousis, C. Z. Patrikakis, “Deep learning empowered wearable-based behavior recognition for search and rescue dogs,” *Sensors*, vol. 22, no. 3, 2022, doi: <https://doi.org/10.3390/s22030993>.

[41] Y. Sangve, Y. Firke, S. Shinde, S. Patil, P. Shinde, P. Mitake, “A comprehensive survey on real-time animal (dog) detection system using artificial intelligence methods,” in *Proceedings of the 4th International Conference on Artificial Intelligence and Smart Energy*, Cham, Switzerland: Springer Nature Switzerland, 2024, pp. 260–275.

[42] O. Akinwehinmi, “Abandoned dogs detection dataset,” 2025. Accessed: Jun. 3, 2025, [Online]. Available: <https://www.kaggle.com/datasets/oluwakemiakinwehinmi/abandoned-dogs-detection>.

[43] O. Akinwehinmi, “Abandoned dog detection - code for reproduction,” 2025. Accessed: Jun. 3, 2025, [Online]. Available: <https://github.com/DrKem/Abandoned-DogDetection/blob/main/Codes>

[44] B. D. Lucas, T. Kanade, “An iterative image registration technique with an application to stereo vision,” in *Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJCAI’81)*, San Francisco, CA, USA, 1981, pp. 674–679.

[45] B. K. Horn, B. G. Schunck, “Determining optical flow,” *Artificial Intelligence*, vol. 17, no. 1, pp. 185–203, 1981, doi: [https://doi.org/10.1016/00043702\(81\)90024-2](https://doi.org/10.1016/00043702(81)90024-2).

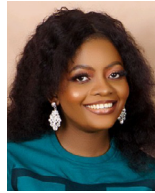
[46] A. Bewley, Z. Ge, L. Ott, F. Ramos, B. Upcroft, “Simple online and realtime tracking,” in *Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, USA, 2016, pp. 3464–3468.

[47] N. Wojke, A. Bewley, D. Paulus, “Simple online and realtime tracking with a deep association metric,” in *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, China, 2017,

pp. 3645–3649.

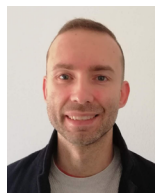
[48] K. Chatfield, A. Zisserman, “Visor: Towards on-the-fly large-scale object category retrieval,” in *Computer Vision – ACCV 2012*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 432–446.

[49] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, A. Zisserman, “The pascal visual object classes (voc) challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010, doi: <https://doi.org/10.1007/s11263-009-0275-4>.



Oluwakemi Akinwehinmi

Oluwakemi Akinwehinmi is currently pursuing a Doctor of Philosophy (PhD) in Computer Science at Universitat de Lleida, Spain (2022–2025). She holds a Master of Science and a Bachelor of Science in Computer Science from the University of Ibadan, Nigeria (2018–2020 and 2012–2016, respectively). Her research interests include artificial intelligence, deep learning, computer vision, generative artificial intelligence, and cloud computing. She is a member and volunteer of Data Science Nigeria, Data Science Africa, and Women in Machine Learning and Artificial Intelligence.



Alberto Tena

Alberto Tena is a lecturer professor at the University of Lleida (UdL), affiliated with the Department of Computer Science and Digital Design, where he specializes in distributed computing. He received a Ph.D. in Engineering and Information Technologies from the University of Lleida in 2022, and holds a Bachelor’s degree in Telecommunications Engineering and a Master’s degree in Applied Telecommunications and Engineering Management, both obtained in 2009 from the Polytechnic University of Catalonia (UPC), Spain. His research focuses on distributed monitoring systems, combining digital signal processing, distributed computing architectures, and machine learning applied to multimodal data.



Francisco Javier Mora

Francisco Javier Mora received a degree in Telecommunications Engineering from the Polytechnic University of Catalonia (UPC) in 1992 and a Ph.D. in Electronic Engineering in 1998. Since 1998, he has been a Senior Researcher and Project Manager at the International Center for Numerical Methods in Engineering (CIMNE). His early career focused on computational electromagnetics and finite element simulation. Currently, he specializes in the digitalization of the AECO industry, focusing on Building Information Modeling (BIM) and expanded reality (XR) applications. His research integrates AR/VR for quality control, monitoring, and high-precision positioning. He plays a leading role in construction tech, bridging the gap between numerical methods and immersive digital tools for industrial innovation.



Francesc Solsona

Francesc Solsona is a full professor at the Universitat de Lleida. His research experience is reflected in four awarded Research Six-Year Terms (sexenios), an h-index of 21 (Google Scholar), and over 1,700 citations. He has published more than 130 journal articles and book chapters, many indexed in the JCR, and has contributed to over 100 national and international conferences. His research focuses on cluster and distributed computing, cloud, fog/edge, and IoT systems, with particular emphasis on the design of algorithms, models, and simulation environments for task scheduling in parallel and distributed architectures. His work is highly interdisciplinary, spanning computer science, medical informatics, epidemiology, artificial intelligence, signal processing, and operations research. He is a member of the Distributed Computing Group (GCD) at the University of Lleida, a consolidated research group accredited by the Government of Catalonia. He has supervised 10 PhD theses, several with international mention and excellence awards, and has participated in more than 30 competitive research and technology transfer projects, including national, European, and industrial initiatives.



Pedro Arnau del Amo

Pedro Arnau del Amo holds a Ph.D. in Physical Oceanography from the Polytechnic University of Catalonia (UPC), Spain. His doctoral thesis on mesoscale marine circulation in the Catalan Sea was awarded First Prize in the Sustainable Development Thesis Competition by the AGBAR Foundation. He currently serves as Chief Climate Resilience Scientist at Findspo. He has led national and international research projects as principal investigator and has been involved in initiatives focused on protecting ecosystem services and integrating information and communication technologies (ICT) into environmental monitoring. His expertise includes remote sensing, geographic information systems (GIS), and machine learning applied to environmental and marine sciences.